

Bases de Données et Système d'Information Géographique : développement des bases de données et des applications de prétraitements

Sylvain Théry, Nicolas Bacq.

UMR 7619 Sisyphe, Boite 123, 4 place Jussieu, 75005 PARIS

thery@ccr.jussieu.fr - bacq@ccr.jussieu.fr

| | |
|--|----|
| Introduction | 1 |
| 1. Inventaire, gestion, mises à jour et acquisition de bases de données spatialisées | 2 |
| 2. Création d'un catalogue des métadonnées selon les normes de gestion de l'information géographique | 5 |
| 3. Développement d'outils de création et de manipulation de bases de données pour leur exploitation et les échanges entre modèles..... | 6 |
| 4. Mise au point de jeux de données pour l'application Senèque 3.1 | 6 |
| 4.1. Extension de Prétraitement de données pour Senèque | 6 |
| 4.2. Développement de jeu de données Senèque 3.1 | 8 |
| Conclusions | 9 |
| Annexes | 10 |
| Bibliographie, références internet | 10 |
| Inventaires des bases de données recensées | 11 |
| Ressources matérielles et logicielles pour les SIG & BD à Jussieu | 12 |

Introduction

Le travail sur les bases de données du PIREN-Seine s'est organisé autour de quatre axes durant l'année 2002 :

- Inventaire, gestion, mise à jours et acquisition de bases de données spatialisées ;
- Développement du catalogue des métadonnées selon les normes de gestion de l'information géographique.
- Développement d'outils de création et de manipulation de bases de données pour leur exploitation et les échanges entre modèles ;
- Mise au point de jeux de données pour l'application Senèque 3.1

1. Inventaire, gestion, mises à jour et acquisition de bases de données spatialisées

L'inventaire effectué au cours de l'année a permis d'identifier les bases de données acquises par les équipes du PIREN au cours des précédentes phases du programme.

Ces bases de données ont été centralisées à Jussieu. Il s'agit des données acquises par les équipes auprès de partenaires extérieurs ou les données mises à dispositions par d'autres programmes de recherche dans le cadre de **licence d'utilisation accordées à l'ensemble du PIREN-Seine**. La description détaillée de ces bases de données recensées et centralisées à Jussieu est présentée en annexe de ce document.

On observe que la plupart de ces bases sont très fragmentaires. Certaines ont été constituées petit à petit, par ajouts successifs de données hétérogènes. Aussi les séries sont souvent incomplètes et limitent leur utilisation par les modèles.

Au de la de cet inventaire, l'identification de données utiles au fonctionnement des modèles mis en œuvre par les équipes a nécessité l'acquisition de nouvelles bases, ou la mise à jour de bases de données existantes auprès de partenaires extérieurs. Ont été obtenues gracieusement ou acquises dans le cadre de convention les bases suivantes :

Vecteurs

- BD Carthage 2.4 (AESN) ;
- BD Cantons (AESN) ;
- BD Ouvrages (AESN) ;
- Extraits de la BD GeoFla Départements (IGN)
- BD HER ;
- BD PRA ;
- BD ZHR ;
- BD Nitrates.
- BD épandages des boues de Step (SIAAP)

Raster

- Couvertures LMDR
- Couverture de scènes satellites SPOT panchromatique à 10 m de résolution sur l'ensemble du bassin versant de la Seine (CNES-USGS, libre de droit)
- Couverture de scènes satellites LandSat multi spectrales (7 canaux) à 30 m de résolution sur l'ensemble du bassin versant de la Seine (NASA-USGS, libre de droit)

Ces deux dernières bases sont en cours de transfert. Elles sont d'un volume substantiel aussi des priorités ont été établies dans l'ordre de leur déploiement à Jussieu.

Base de donnée non-spatialisée :

- Inventaire du RGA 2000 (Agreste – Ministère français de l'Agriculture)

Les nouvelles bases sont opérationnelles en vue de leur exploitation cartographique et/ou sous logiciels de SIG.

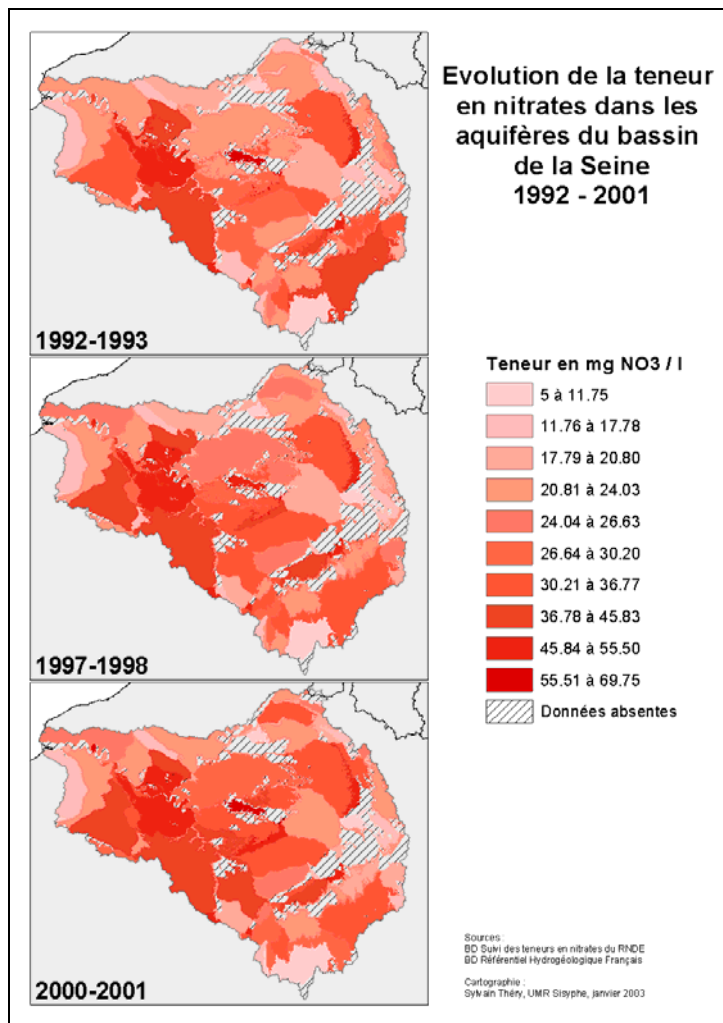


Figure 1 :

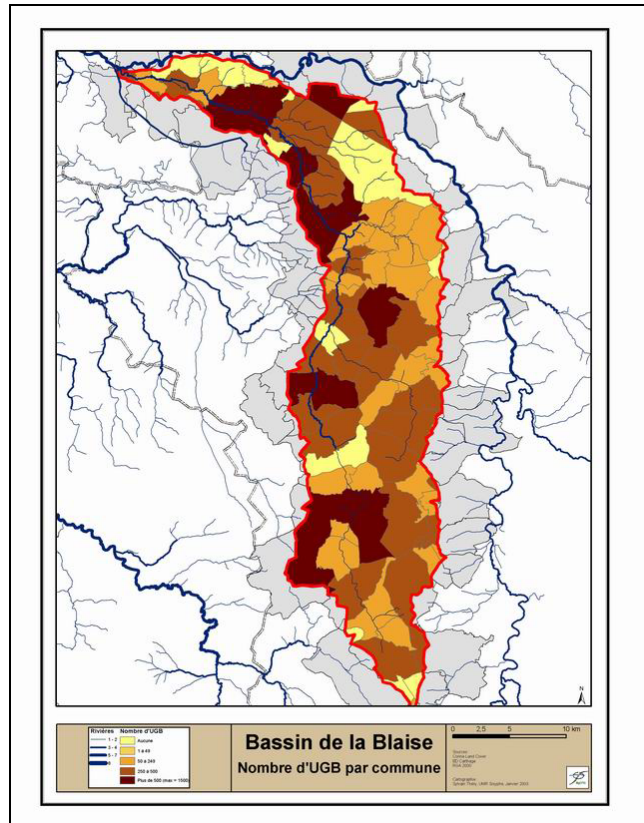
Exemple d'explorations cartographiques des bases de données nouvellement acquises : utilisation des bases de données géographique du RNDE et du Référentiel Hydrographique français.

Toutes ces bases ont été acquises et/ou obtenues sous licence PIREN-Seine, permettant ainsi leur diffusion sans restrictions à l'intérieur du programme. Certaines ont été prétraitées de manière à être immédiatement exploitables sous Arcview 3.x et Arc-Info 8.1. Ces prétraitements concernent :

- la vérification du géoréférencement des bases (changement de système de référence spatial si nécessaire, acquisition / création d'un référent spatial) ;
- la validation / correction de la topologie ;
- la validation de la structure physique des tables attributaires (transformation si nécessaire des structures pour les mettre en conformité avec les structures utilisées au PIREN) ;
- l'ajout de données attributaires propres à l'utilisation faite de ces bases au sein des programmes de recherche.

Figure 2 :

La résolution de certaines bases permet le traitement des données géographiques à des échelles très fines : exemple d'exploitation des données du RGA 2000 rapportées à l'échelle communale



D'autres bases de données ont été localisées et devraient être prochainement déployées :

- Corinne LandCover 2 (IFEN);
- BD Carthage 3 (AESN)
- Base de données de l'Aube (CG Aube)
- Scan 25 (AESN)

Un autre aspect de la centralisation des bases de données du PIREN concerne les bases créées par les équipes du PIREN. Certaines d'entre elles ont été intégrées à la base générale du PIREN :

- Simulation MODCOU sur le bassin de la Marne
- Typologie des zones humides ripariennes.
- Scan des cartes topographiques (1/25000), géologiques (1/25000) et hydrogéologiques (1/50000) de la Marne, de l'Eure et du bassin de la Seine.

Mais la plupart des données élaborées par les équipes ou acquises individuellement n'ont pas été centralisées. Leur spécificité minore l'intérêt de leur transfert au sein de la base de données commune des équipes. Par ailleurs, leurs conditions d'acquisition limite souvent leur champ de diffusion.

2. Création d'un catalogue des métadonnées selon les normes de gestion de l'information géographique

La multiplication des bases entraîne la nécessité d'un catalogage non seulement des données, mais également des métadonnées, informations décrivant les bases sous différents aspects (géoréférencement, structure, méthodologie, intervenants, etc.).

Ceci est d'autant plus nécessaire que la confrontation de données émanant de sources différentes et regroupées au sein de la base du PIREN montre des incohérences. Ceci peut s'expliquer par l'absence de métadonnées qui préciseraient les méthodologies, sources, et éventuellement prétraitements effectués sur ces bases. De ce fait, leur intégration est longue et incertaine.

Les préoccupations concernant la gestion des catalogues de métadonnées ne sont pas nouvelles aussi un rapide état de l'art dans le domaine a permis de suivre plusieurs pistes méthodologiques en vue de la création du catalogue des bases de données du PIREN :

- Utilisation de la norme CNIG à travers l'utilisation du logiciel REPORT v.2 développé par le CERTU.
- Utilisation des recommandations du Dublin Core Metadata Initiative (DCMI).
- Utilisation de la norme CSDGM (Content Standard for Digital Geospatial Metadata) du FGDC.

Les deux premières méthodologies n'ont pas été retenues, soit parce qu'elles ne sont pas assez structurées (comité consultatifs ne débouchant pas sur la création d'outils opérationnels), soit parce que leurs tests d'application n'ont pas été concluants (instabilité du module de consultation en ligne des outils du CERTU sur la machine hébergeant le site web du PIREN). Par ailleurs, aucune d'entre elles ne permet de lier de manière informatique les modèles et les données.

Les bases de données sont donc en cours d'indexation et de catalogages selon les normes FGDC-STD-001-1998, à l'aide d'outils ad hoc.

Le choix de cette norme et de cette méthodologie a été dicté par plusieurs aspects :

- Cette norme est directement compatible avec la gestion des modèles sous Arc Info 8 et indirectement avec Arcview 3.x (nécessité d'utiliser une extension tierce, libre de droit).
- Le stockage des informations se fait sous la forme de fichiers ascii de type XML, permettant ainsi des échanges aisés avec différents formats de gestion de base. De plus ce format offre une grande souplesse dans la définition des éléments catalogués.
- Ces facilités d'échanges permettent une mise en ligne semi-automatique des métadonnées (on peut voir sur le site web du PIREN les descriptions des bases de données des bases déjà cataloguées).

Toutes les nouvelles bases sont donc cataloguées au fil de leur acquisition tandis qu'un travail de collecte des métadonnées est en cours. Le mode de catalogage est encore susceptible d'évoluer, de même que son mode de mise en ligne sur le site web du PIREN.

Les extractions de données les plus fréquemment demandées par les équipes sont également mises en ligne, au fil des demandes. Il ne s'agit donc pas de la mise en ligne d'une base spatialisée consultable directement par l'utilisateur, mais d'une série d'extraits des bases communes à tout les programmes. Ces extraits ne concernent, pour des raisons de droit, que des bases élaborées par le programme.

La réalisation d'une base de données géographique en ligne n'est sans doute pas à considérer comme une priorité, vu les problèmes techniques, et surtout juridiques qu'elle engendrerait (propriété intellectuelle des données, droits d'exploitation en fonction des licences et conventions), outre le coût substantiel induit (licence ArcIMS, ou autre). Par ailleurs, la relative nouveauté des technologies de mise en ligne n'a pas encore permis la définition d'un format standard d'exploitation en ligne des bases de données géographiques. Des travaux et réflexions sont cependant menés en vue de la définition de tels standards, et le PIREN s'est impliqué au sein de deux groupes :

- Intercarto : groupe de réflexion menée par le réseau francophone ‘Géomatique’ ;
- GRASS (Geographic Resources Analysis Support System) : groupe international de réflexion et de développement d’outils de manipulation de l’information géographique

3. Développement d’outils de création et de manipulation de bases de données pour leur exploitation et les échanges entre modèles

Les prétraitements des bases en vue de leur exploitation au sein du programme constituent une part substantielle de l’activité, aussi un objectif important constitue l’automatisation de ces tâches. Plusieurs logiciels ont été acquis à ces fins mais le caractère spécifique de la plupart des traitements entraîne la programmation de routines de traitement automatisé.

Ces développements sont programmés en ayant recours aux logiciels de SIG Arcview et Arc-Info. Ils se font en langage Avenue et en langage Visual Basic. L’objectif à terme est d’effectuer l’ensemble de la chaîne des prétraitements sous Arc-Info 8 (obsolescence d’Arcview, intégration des normes récentes de programmation et gestion de BD dans sous Arc-Info, interface user-friendly).

Certaines des routines développées servent à créer des passerelles entre les différents modèles et/ou bases de données.

Des applications d’échanges de données entre modèles ont ainsi été développées :

- Automatisation des échanges de données entre Arpège et les données du LMDR à travers les formats NetCDF et le maillage raster Arc-Info (développement en AML)
- Echange de données entre STICS-Modcou et RiverStrahler (format Seneque 3.1, développement en Avenue)
- Intégration de données externes en ascii (VB)

Au-delà de ces développements ponctuels, le principal projet est la définition d’un réseau de sections mouillées de Prose compatible avec les réseaux tels que représentés dans Seneque 3.1. Ceci montre l’intérêt de la définition d’un réseau hydrographique de surface commun à Prose et à Seneque. Un tel réseau devrait être doté de plusieurs représentations en fonction de l’échelle et du modèle utilisés. Un développement de la base de données dans cette direction impliquerait la collaboration aux travaux déjà réalisés dans ce domaine (réalisations de l’EPFL, notamment).

D’autres routines sont utilisées pour agréger ou désagréger les informations en fonctions des échelles. L’automatisation des extractions à partir des bases de données entraîne également la programmation de routines sous logiciels de traitement SIG. Les développements en cours portent essentiellement sur :

- une application de type ‘système expert’ concernant les pollutions en métaux lourds ;
- des traitements concernant des données d’épandage des boues des stations d’épuration.

4. Mise au point de jeux de données pour l’application Seneque 3.1

4.1. Extension de Prétraitement de données pour Seneque

Devant l’importance des bases de données nécessaires au fonctionnement du modèle RiverStrahler, une application spécifique a été développée : EXPRES 1.0 (EXtension de PREtraitement pour Seneque).

Cette application est développée en langage Avenue et vient donc se greffer sur le moteur et l'interface d'Arcview 3.x et de Spatial Analyst. Elle est destinée à générer des jeux de données pour leur exploitation par le logiciel Seneque 3.1. La définition d'un format générique des données d'entrée de l'applicatif permet d'utiliser des données issues de sources hétérogènes mais impliquent de nombreux prétraitements. La plupart sont automatisés dans cette extension :

- Validation ou création de la géométrie des réseaux d'axes et bassins ;
- Création et/ou mise en conformité des couvertures géographiques avec la structure du modèle de données de Seneque ;
- Manipulation et validation des données tabulaires nécessaires à l'applicatif ;
- Création de la codification Pfsafstetter-Verdin ;
- Création de couvertures spécifiques en vue de modélisation prospectives ;
- Génération des fichiers clés pour l'occupation du sol, l'hydrogéologie et les apports diffus ;
- Gestion des métadonnées ;
- Etc.

Pour une description complète des fonctionnalités de prétraitements, on peut se référer à la documentation de l'extension.

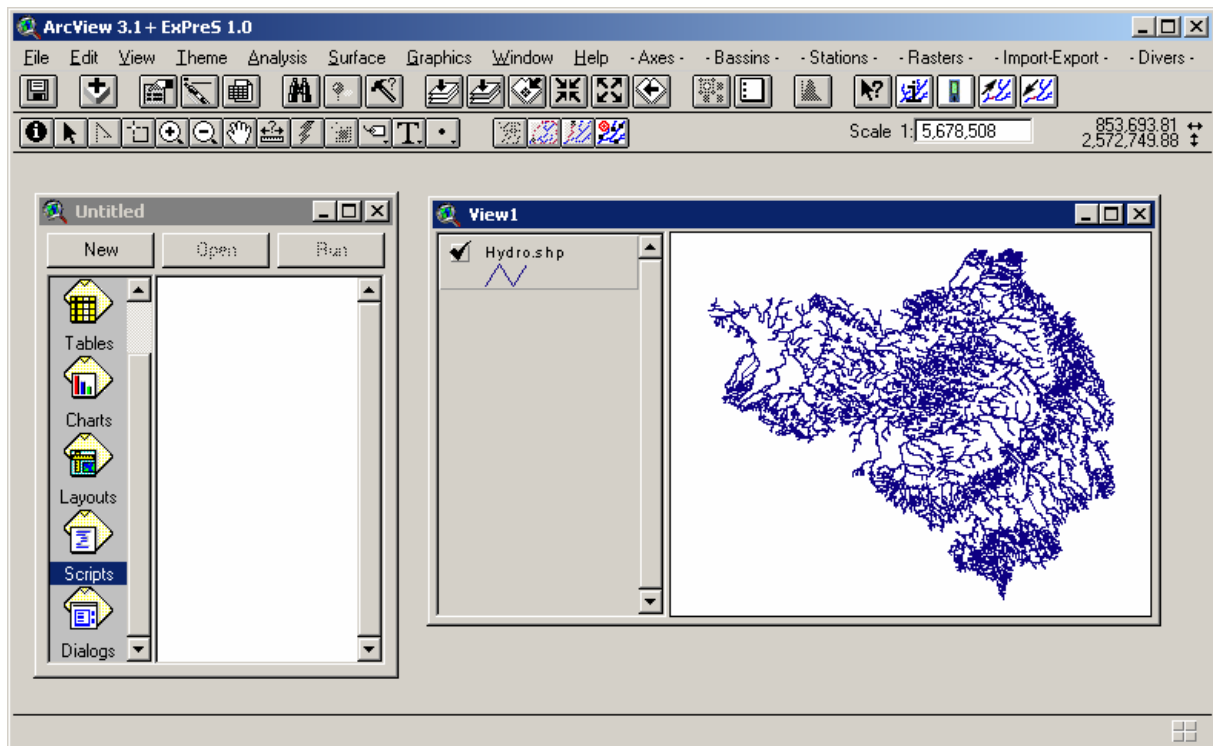


Figure 3 : ExPreS 1.0 s'intègre dans l'interface d'ArcView. Les menus et icônes de l'extension permettent de lancer différents traitements sur les données en vue de leur utilisation dans l'application Seneque

L'extension est déployée en phase de test à l'AESN-Nanterre, Compiègne et Reims, ainsi qu'à l'Ecole des Mines de Paris. La dernière version développée est mise en ligne sur le site du PIREN.

4.2. Développement de jeu de données Seneque 3.1

Après le développement du modèle RiverStrahler sur la Marne, l'Oise est le deuxième territoire dont l'hydrosystème fait l'objet de modélisation avec Seneque. L'établissement du jeu de données s'est fait tout d'abord pour l'application Seneque 1.3, afin d'être utilisés dans le cadre des travaux de l'AESN concernant la directive cadre européenne.

Les évolutions de Seneque 3.1 ont entraîné un changement de la structure des bases de données, pour aboutir à un nouveau jeu de données en vue de sa présentation au colloque du PIREN de février 2003.

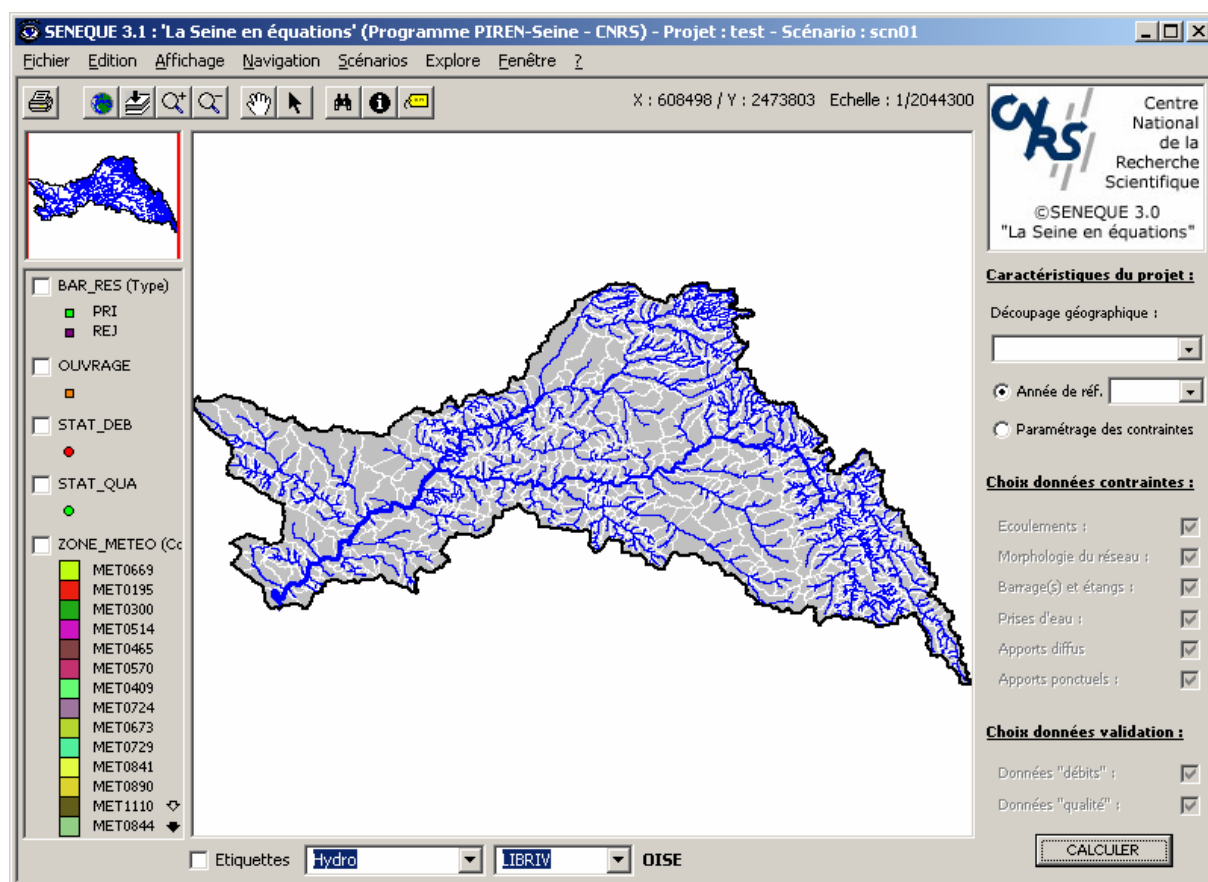


Figure 4 : l'application Seneque 3.1 sur le bassin de l'Oise

D'autres nouveaux jeux de données doivent être développés :

- Jeu de données sur le bassin de la Vesle
- Jeu de données pour l'intégralité du bassin de la Seine (échéance : juin 2003)
- Jeu de données pour les bassins côtiers bas-normand (échéance : été 2003)

Afin de faciliter le regroupement des données nécessaires à l'applicatif, l'AESN envisage de créer un comité de pilotage de l'utilisation de Seneque dans ces représentations.

Conclusions

La création de nouveaux jeux de données pour l'application Seneque est donc en cours. Elle entraîne l'évolution en parallèle de l'extension de prétraitements. Les fonctionnalités à venir devront intégrer des outils de délimitation de bassins plus efficaces, afin de pallier l'absence de ce référent spatial dans certaines zones.

Ceci participe également au rapprochement entre les différentes représentations de l'hydrosystème selon les modèles : les réflexions menées autour du modèle de données de PROSE s'axent vers la définition de formats de données proches de ceux utilisés par Seneque.

Les restructuration des bases et l'acquisition de nouvelles données complètent le catalogue des bases de données. Pour les données anciennes, des métadonnées sont établies rétroactivement de manière aussi complète que possible. Ces métadonnées sont mises en ligne au fil de leur élaboration.

Cette centralisation des bases pose de nouveaux problèmes notamment en ce qui concerne leur conservation. La sécurité des bases et leur meilleure exploitation nécessite de nouveaux dispositifs de sauvegardes de masse afin d'éviter la perte accidentelle de données, et de permettre l'archivage des jeux de données au fil de leur évolution.

La mise en ligne des données elles-mêmes implique une réflexion commune autour des questions que la mise en ligne d'Information Géographique : problèmes juridiques liés aux droits de propriété sur les données, contraintes de l'infrastructure technique, etc.

Le développement de l'application métaux est également à l'ordre du jour. L'achat d'une seconde licence Arc-Info permet d'être opérationnel et de lancer ce nouveau projet.

Par ailleurs, les nouvelles bases devraient faire l'objet d'explorations cartographiques et de prétraitements pour leur utilisation par les modèles.

Annexes

Bibliographie, références internet

Métadonnées

Normes DCMI : <http://dublincore.org/>

Normes CSDGM : <http://www.fgdc.gov/metadata/constan.html>

Pôle géomatique du CERTU : http://www.certu.fr/technologies/pole_geomatique/

Métadonnées sous ArcInfo : <http://www.esri.com/metadata/>

Bases à multi représentations et multi échelles

SPACCAPIETRA S., PARENT C., VANGENOT C., "GIS Databases: From Multiscale to MultiRepresentation" in *Proceedings of the International Workshop on Emerging Technologies for Geo-Based Applications*, May 22-25, 2000, Ascona, Switzerland, EPFL-DI-LBD Publisher

SPACCAPIETRA S. et al, *The MurMur Data Modeling Approach*, 62 p. EPFL, 2000.

Structuration des bases de données géographiques

DATE (C.J.), *Introduction aux bases de données*, 926 p., Vuibert informatique, Paris, janvier 2001.

GARDARIN (G.), *Bases de données : objet et relationnel*, 788 p. Eyrolles, Paris, 2001

SCHOTT (M.), *SGBD géographiques : spécificités*, 185 p., International Thomson publishing France, Paris, 1996.

STEINBERG (J.), *Cartographie télédétection systèmes d'information géographique*, Armand Colin, Paris, janvier 2002

Mise en ligne et serveur de bases de données

CLOUX (P.-Y.) et al., *Architectures client-serveur Internet et intranet, De CGI aux EJB*, Interédition, Paris, 215 p., juin 2000.

CELKO (J.), *SQL avancé : programmation et techniques avancées*, Vuibert informatique, Paris, 488 p., août 2000.

PLEWE (B.), *GIS Online: Information Retrieval, Mapping, and the Internet*, OnWordPress, 400 p., 1997.

Applications hydrologiques et SIG

BRUNSTEIN , *Descriptif des bases de données géoréférencées au PIREN*, document PIREN-Seine, 1999.

RUELLAND D., BILLEN G. *Applicatif Seneque 3.0 - Notice de développement*. document PIREN-Seine, 2002.

M. MEYBECK, Y. GUEGUEN, H. DÜRR, C. GROSOBOIS, L. LACHARTRE, N. BACQ, S. THERY, A. HOROWITZ, *Vers un système expert pour réaliser le bilan de la contamination métallique du réseau hydrographique*.

A. ROY DE LACHAISE, M. BLANCHARD, MJ TEIL, D. DURIEZ, A. BRANTHOMME, S. THERY, J-M. MOUCHEL, M. CHEVREUIL, *Valorisation agricole des boues urbaines et industrielles : impact à l'échelle des bassins versants*, Rapports annuel PIREN-2003

N. BACQ, *Apports de la modélisation dans les scénarios de la directive cadre sur l'eau*, Mémoire de DESS, 2002.

STRAHLER A-N. *Quantitative analysis of watershed geomorphology*. Washington : Geophys. Union Trans. 38, pp. 913-920, 1957.

VERDIN K. *A System for Topologically Coding Global Drainage Basins and Stream Networks*, US Geological Survey, 2001

Inventaires des bases de données recensées

Les bases inventoriés et mises en forme sont, soit des bases développées ou mises en forme au PIREN, soit des bases acquises auprès de partenaires extérieurs.

- Réseau hydrographique (PIREN)
- Bassins versants (PIREN)

Couvertures des axes et bassins développées en interne.

- Couverture Corridor Fluvial (PIREN)

Couverture développé en interne.

- BD PRA

Base de données développée par l'INRA Mirecourt.

- BD Population

Base de données développée au laboratoire de démographie humaine (LDH).

- Base qualité (AESN PIREN)

On dispose d'une couverture de la localisation de 1754 sites du réseau de mesure de la qualité de l'eau. Les sites sont diversement renseignés.

- Couverture hydrogéologique

Vectorisation des cartes d'Albient au 1/500000è.

Soit des bases acquises auprès de partenaires extérieurs :

- Base de Données Géographique des Sols de France.

La Base de Données Géographiques des Sols de France au 1/1 000 000 est l'un des composants de la Base de Données Géographique des Sols d'Europe au 1/1 000 000 (SGBDE, version 3.2.8.0, 10/09/1998). Les méta-données de la base française suivent le formalisme de la base européenne.

- Base Corine Landcover

Conduit en France par l'IFEN, CORINE Land Cover est une base de données géographiques de l'occupation des terres réalisée dans le cadre du programme européen CORINE (Coordination de l'Information sur l'Environnement) mené par l'Agence Européenne pour l'Environnement (<http://www.eea.eu.int>).

Etendue : Bassin de la Seine

- Recensement Général de l'Agriculture

Inventaire partiel de 1988.

- Banque "Hydro"

Nous disposons de données quotidiennes et mensuelles de 362 stations de mesure du réseau du bassin de la Seine.

- Banque "Pluvio"
10 stations : longue période
10 stations : période 1995-1998
- Données d'évapotranspiration
12 stations synoptiques.
période 1995-1998.
- Banque piézométrique
La base contient les mesures quotidiennes du niveau des nappes à 226 piézomètres.
- BD Carthage
AESN / IGN. Etendu : bassin de la Seine.
- Zones hydrologiques
Limites des zones hydrologiques.
Les zones correspondent à des bassins versants d'une superficie de l'ordre de 500 km².
Chaque zone est décrite par un code hydrologique (correspond aux 4 premiers caractères du code hydrologique des segments du réseau).
Etendue : Bassin versant de la Seine.
- Découpages administratif
Départements et communes du bassin de la Seine.
- Digital Chart of the World
- Base Gtopo30
La résolution de est 30'' d'arc. La résolution kilométrique varie donc avec la latitude.
Le système de coordonnées original est en degrés décimaux sur l'ellipsoïde WGS84.
L'altitude est exprimée en mètres depuis le niveau moyen de la mer.
Etendue : monde.
- Base Hydro1k
HYDRO1k, est une base de données géographiques fournissant une couverture à l'échelle globale de données généralement utilisées pour des applications hydrologiques.
Développée par le centre EROS Data de l'U.S. Geological Survey's (EROS Data Center) en collaboration avec Programme des Nations Unies pour le Développement/Global Resource Information Database (UNEP/GRID).
Etendue : monde.

Ressources matérielles et logicielles pour les SIG & BD à Jussieu

Ordinateurs :

- un Pentium 400 mhz, 128 mo de RAM, disque dur de 4 giga-octets.
- un Pentium 400 mhz, 256 mo de RAM, disque dur de 4 giga-octets.
- un Pentium 2.4 ghz, 512 mo de RAM, disque dur de 60 giga-octets.
- 1 imprimante jet d'encre couleur A4.

Logiciels :

- Arview 3.1, Spatial Analyst et Network Analyst (SIG).
- Deux licences Arc Info 8.1 Labkit (SIG).
- Dreamweaver (XML et HTML)
- MS-Access (SGBD)
- Visual Basic 6 (Programmation)
- Java developpemet kit (JDK 1.2)
- MapObjects 2 (ActiveX)
- InnoSetup (Déploiement d'application)

Mise en ligne :

- Serveur Apache, php et mySQL sous Windows 2000 de l'UMR Sisyphe.